44° Convegno Internazionale
dei Docenti delle Discipline della Rappresentazione
Congresso della Unione Italiana per il Disegno

Disegno e intelligenza artificiale. Enunciati teorici e prassi sperimentale per una poiesi condivisa

Francesca Condorelli
Alessandro Luigini
Giuseppe Nicastro
Barbara Tramelli

Abstract

La pervasività delle tecnologie digitali sta favorendo un rapido sviluppo dell'insieme delle tecnologie che ricadono all'interno della denominazione Intalligenza Artificale (AI), e come molte delle innovazioni tecnologiche digitali hanno un impatto notevole sulla produzione di immagini e disegni. Studiare come la creatività si interfacci con un sistema AI dotato di notevole autonomia nella produzione di una immagine è fondamentale per comprendere come l'autore debba operare per governare lo strumento e guidarne la poiesi.
Il contributo propone una contestualizzazione di carattere teorico del fenomeno delle applicazioni text-to-image nel panorama della storia dell'arte e della produzione di immagini, strettamente collegata con una esplicitazione del funzionamento di tali applicazioni, per poi verificarne le potenzialità nel campo della narrazione visuale.

Parole chiave
Intelligenza artificiale, narrazione visuale, poiesi

Alcuni esempi di immagini generate con l'applicativo *prompt-to-image* Midjourney.

doi.org/10.3280/oa-1016-c426

Introduzione

La pervasività delle tecnologie digitali sta accelerando lo sviluppo dell'insieme delle tecnologie che ricadono all'interno della denominazione di Intelligenza Artificiale (AI), con applicazioni *web-based* accessibili da qualunque dispositivo informatico, inclusi i dispositivi mobili. La generazione di testi verbali e visivi è diventata possibile grazie ad AI come ChatGPT e Midjourney e i risultati sono convincenti, seppur ancora in via di ulteriore sviluppo.
Il paper che qui si presenta parte da l'enunciazione teorica del contesto in cui queste applicazioni si sviluppano, prosegue con la descrizione del processo con cui le immagini vengono generate e termina con la presentazione di una sperimentazione in ambito didattico utilizzata come verifica pragmatica delle tesi precedentemente enunciate.

Immagini dalle parole

I modelli di intelligenza artificiale in grado di creare immagini da *prompt* (i cosiddetti generatori *promt-to-image*) come Midjourney, DALL.E e Stable Diffusion sono diventati strumenti sempre più apprezzati da artisti e designers, e come spesso accade in molti campi che concernono le *digital humanities*, la teoria arriva necessariamente in un secondo tempo, sulla base dell'esperienza pratica. Il fenomeno dei sistemi di *machine learning* che creano immagini basandosi su collezioni di dati è ormai noto, e molte delle questioni relative all'utilizzo e all'usabilità di questi sistemi (ad esempio, l'annosa e tutt'altro che risolta questione dei copyright) sono già state affrontate in vari modi, con varie possibili soluzioni [1]. La loro crescente popolarità ci interroga sulla necessità di fornire come studiosi una cornice e un contesto teorico nel campo in costante evoluzione del *visual imaging*.
Tali strumenti consentono una rappresentazione figurativa a partire dalla scrittura di un *prompt* descrittivo, e i risultati visivi possono essere sorprendentemente dettagliati [2]. Per quanto riguarda la qualità delle immagini, Midjourney per esempio tende ad avere uno stile riconoscibile tra gli altri per la quantità e la qualità dei dettagli forniti, mentre altri (DALL.E in particolare) si caratterizzano per il realismo nella illuminazione delle scene ma anche per la tendenza a dimenticare le parole chiave quando i *prompt* diventano troppo dettagliati [Sambucci 2022]. Lo stile è abbastanza coerente nelle diverse immagini fornite dai modelli, come ha affermato il fondatore David Holz: "il nostro sforzo si concentra sul rendere le immagini esteticamente belle" e questa è una caratteristica fondante delle immagini create dall'IA, generalmente accattivanti alla vista, si potrebbero definire 'piacevoli' per gli occhi di chi guarda [Vincent 2022]. La qualità di queste immagini può essere misurata considerando sia il livello di realismo dell'output visivo, sia il numero di dettagli che rappresentano e interpretano il *prompt* linguistico inserito. Quest'ultimo punto è di particolare interesse, poiché comporta il bisogno di un'analisi più approfondita del rapporto tra interazione umana e intelligenza artificiale, che non può limitarsi all'assunto causale 'scrivo una parola, ottengo un'immagine'. Il processo di creazione di queste immagini è molto più complesso, e ciò si dimostra soprattutto quando il sistema non riesce a generare il risultato atteso, come illustrano diversi tentativi 'fallimentari' [Carlson 2022].
In effetti anche Midjourney occasionalmente omette qualche dettaglio inserito nei *prompt*, mantenendo comunque un livello convincente di *imaging* (fig. 1), mentre i due problemi principali riscontrati durante l'uso dell'algoritmo sono una generale difficoltà congenita nella gestione degli aspetti tipografici (fig. 2), e l'incapacità quasi difficilmente comprensibile – se paragonata al realismo di volti e altri dettagli generati – nella generazione di immagini di mani umane, sia per forma che per correttezza anatomica (fig. 3).
È noto che l'intelligenza artificiale impara e migliora per tentativi, e soprattutto impara dai suoi errori, che vengono forniti sulla base dei dati che inseriamo. Più dati inseriamo, più scelte compiamo, più i risultati sono accurati. Questi ultimi possono sorprenderci, in quanto utilizzando varie parole chiave possiamo ottenere 'molto più' o 'molto meno' del previsto, in un modo non dissimile alla serendipità con cui un'immagine creata ha successo o meno in rete. Midjourney, ad esempio, offre solitamente come risultato visivo immagini oniriche con uno

Fig. 1. #*Fallingwater house by Frank Lloyd Wright on mars with a blue sunset and a flying horse*. Evidentemente nella elaborazione di Midjourney manca il cavallo alato, pur avendo generato una immagine di grande impatto visivo e con una propria coerenza semantica rispetto al *prompt*.

Fig. 2. #*happy new year, 2023, Yayoi Kusama style*. In questa quaterna di immagini generate in Midjourney è evidente la difficoltà di gestione dei caratteri tipografici: l'AI comprende che è necessario inserire un numero nell'immagine, probabilmente dalla prima parte del *prompt 'happy new year'* ma non riesce a inserire il numero giusto, nonostante sia presente nel *prompt* e sia stato isolato dalla punteggiatura. Lo stile dell'artista, invece, è correttamente interpretato.

Fig. 3. A sinistra, #*drawing resembling a black and white engraving of a hands that old a pencil*. A destra, #*black and white illustration of a hand*. Queste serie di immagini sono due delle molte serie elaborate sul tema delle mani che disegnano, e i risultati, come si vede chiaramente, sono molto affascinanti ma perdono completamente i riferimenti della correttezza anatomica e della verosimiglianza che rendono le immagini di Midjourney così efficaci.

stile neogotico molto accattivante, di facile leggibilità per l'occhio e intuitivo da interpretare, il che rende il sistema particolarmente adatto a illustrare racconti e fiabe [3]. Il ruolo dell'intelligenza artificiale è ovviamente centrale nel processo, ma il ruolo dell'utente che inserisce le parole chiave nel *prompt* non è di minor rilevanza. Come affermato in un recente articolo di Ziv Epstein: "Nessuna IA agisce da sola, completamente separata dall'influenza degli umani. Anche l'opera d'arte *Edmond de Belamy*, che si diceva fosse 'interamente realizzata da un algoritmo', era il risultato della creatività, del duro lavoro e delle decisioni di numerosi attori umani. Quando un sistema di intelligenza artificiale realizza qualcosa di eccezionale o causa un problema serio, in che modo viene ripartita la responsabilità agli umani che lo circondano?" [Epstein et al. 2020].
Nel caso di Midjourney, siamo consapevoli del ruolo essenziale svolto dalle *keywords*, e quindi dagli utenti. Il processo di creazione artistica sulla base dell'analisi di materiale artistico fornito da banche dati (in questo caso milioni di immagini) e il combinare questo materiale di partenza senza *authority control* non è certamente nuovo nella storia dell'arte, dato che sappiamo come esperimenti creativi del genere siano stati realizzati ben prima dell'arte creata dal computer.
L'ingerenza della tecnica e della tecnologia nella produzione artistica è stata da sempre un elemento importante per distinguere stili e periodi, intenzioni di singoli autori o sistemi di collaborazione tra più autori. Dal momento in cui una determinata tecnologia è utilizzabile da più persone, non necessariamente contemporaneamente e neppure allo stesso modo, la questione della autorialità – ampiamente dibattuta nella storia dell'arte – è e resta aperta. Quando nel 1975 Brian Eno pubblica il concept album *Discreet Music*, in cui sperimenta diverse tecniche di generazione del suono alternative agli strumenti musicali tradizionali, dichiara il suo programma poietico esplicitamente sulla cover dell'album: "ho gravitato alla volta di situazioni e sistemi che, una volta divenuti operativi, potessero creare musica con interventi minimi o addirittura senza interventi da parte mia. Cioè, io tendo al ruolo di programmatore e pianificatore e in seguito a quello di ascoltatore dei risultati". Nel campo delle espressioni estetiche visuali si potrebbero citare innumerevoli esperienze analoghe, ma per necessaria brevità sarà sufficiente ricordare che fin dagli anni '60 esperienze come quella di Desmond Paul Henry e delle sue *Drawing Machine* – dispositivi elettro-meccanici capaci di tracciare segni su un foglio seguendo schemi matematici – il tentativo di perseguire un processo poietico automatizzato si susseguono con risultati alterni. È così per l'*Autoillustrator* di Adrian Ward o il *Poietic Generator* di Oliver Auber che indagano le possibilità di una poiesi collaborativa e non automatizzata: sono gli utenti-autori che interagendo generano configurazioni visuali, realizzando così una poiesi condivisa che però è resa possibile dalla ideazione dell'autore del sistema. Potremmo proseguire il nostro excursus citando i pionieri dell'architettura digitale che negli anni '90 del Novecento hanno indagato le potenzialità
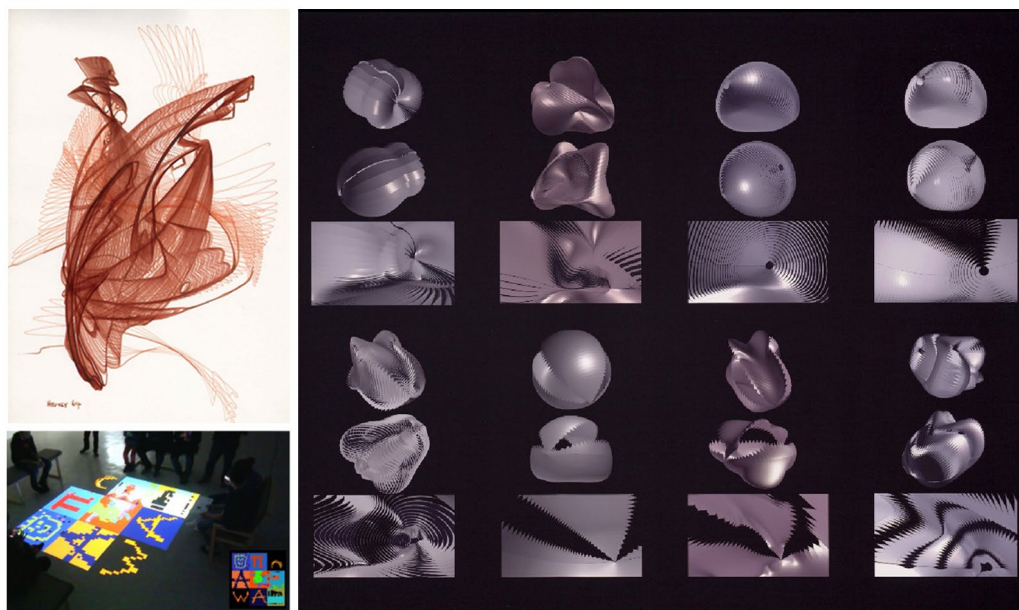
Fig. 4. In alto a sinistra, Desmond Paul Henry, *Untitled*, 1964, <desmondhenry.com>. In basso, Olivier Auber, *Poietic Generator*, 1986-2017, <narrabase.net/auber.html>. A destra, Greg Lynn, *Embryologic House*, 1998, <sfmoma.org/artist/Greg_Lynn>.

dei modellatori avanzati e dei sistemi generativi per percorsi poietici non tradizionali, come Marcos Novak, Greg Lynn, Karl Chu o gli Asymptote (fig. 4).
In quest'ottica, allora, non è l'automazione, il sistema di generazione della immagine distinta dalla azione diretta dell'autore a marcare la differenza tra i sistemi *AI prompt-to-image* e i dispositivi tecnici e tecnologici precedenti. Tra tutte le differenze poniamo l'accento su quella probabilmente più evidente: tutte le opere citate in precedenza, ugualmente autogenerate, sono espressioni estetiche astratte, mentre Midjourney o Dall-E ci consentono di accedere a un immaginario precedentemente inaccessibile e perfettamente fotorealistico. Lo stupore che nell'utente generico produce il primo approccio a questi sistemi è dato normalmente dal realismo, dalla coerenza anatomica e fisiognomica, sia antropomorfa che zoomorfa, e della illuminazione, che dimostrano come questi sistemi siano in grado di concepire il rapporto tra le parti di una immagine rappresentandone la tridimensionalità senza simularla. Il realismo convincente apre la strada a molte applicazioni di *imaging* e le possibilità di applicazione sono molto ampie. Sono già diversi i casi di contest vinti da artisti che hanno presentato un'opera realizzata in Midjourney, come ad esempio Jason Allen, o di creativi che hanno realizzato scenografie, fondali o intere serie di *characters* in Midjourney, come il videogame *Bear Verse* di Eugene Kitkin, ma l'autorialità per ora non sembra essere messa in dubbio.

I sistemi di *prompt-to-image* basati sull'Intelligenza Artificiale: panoramica generale

La nascita delle soluzioni di IA capaci di generare immagini a partire da *prompt* testuali non è da attribuirsi ad un unico filone di studi quanto, piuttosto, alla convergenza di ricerche differenti che hanno reso possibile la nascita di applicativi quali Midjourney, Dall-E, etc [Muppalla et al. 2022].
Il processo consiste di tre passaggi (fig. 5): prima di tutto, un comando di testo viene inviato a un *text encoder* (codificatore testuale) che è stato addestrato per decifrarlo e attribuire a esso dei valori numerici. Il secondo passaggio prevede invece l'intervento di un modello chiamato *prior*, che associa la codifica del testo a una codifica dell'immagine corrispondente, acquisendo le informazioni semantiche del comando testuale. Infine, un *image decoder* genera un'immagine che è una manifestazione visuale delle stesse informazioni semantiche. Relativamente al modo in cui le immagini prendono forma questa fase del processo è correlata, da un punto di vista tecnico, alla possibilità di destrutturare una serie organizzata di specifici dati ed applicare poi un processo inverso che sia in grado di riconfigurare i dati
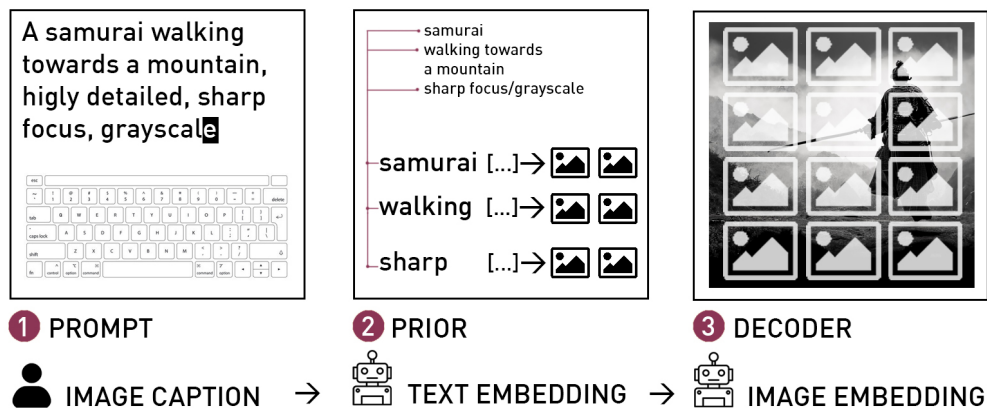
Fig. 5. *Workflow* sintetico di un sistema *prompt-to-image*.

stessi in una composizione quanto più vicina a quella originaria [Sohl-Dickstein et al. 2015]. Rapportandoci alle immagini digitali, si tratta quindi di alterarne e alterare i pixel che la costituiscono, la cui precisa posizione all'interno dello spazio digitale dà vita a ciò che vediamo (fig. 6a), introducendo una quantità nota di rumore (ovvero una quantità di pixel posizionati in maniera casuale) che altera l'immagine stessa (figg. 6b,6c). Lo stadio successivo di questo processo prevede pertanto che, agendo all'inverso (figg. 6d-6f), si possa ricostruire la posizione originaria dei pixel determinando quindi una nuova immagine simile a quella iniziale [Rombach et al. 2022; Ho et al. 2020].
Reiterando questo processo su grandi dataset è quindi possibile generare un modello predittivo che, in maniera autonoma, sia in grado di comporre immagini riordinando una quantità di pixel non strutturati.
Quanto descritto finora ci permette dunque di generare immagini in modo casuale: il passo successivo è pertanto rappresentato dalla possibilità di inserire, all'interno di questa catena, una serie di istruzioni che possano guidare il processo di generazione verso una direzione prestabilita e non casuale.
La possibilità di poter insegnare ad un calcolatore come discriminare e classificare il contenuto di un'immagine è certamente uno dei campi di ricerca che più ha interessato la comunità scientifica che si occupa di *Machine Learning* e Intelligenza Artificiale. Generare un'immagine a partire da un *prompt* testuale comporta infatti che la macchina sia innanzitutto in grado di interpretare un oggetto grafico (una foto, un disegno, etc.) ed associarvi un *tag* testuale che ne descriva il contenuto [Alamar 2022]. Anche in questo caso, le fasi di *training* prevedono l'analisi di *dataset* costituiti da milioni di immagini associate alle rispettive *caption* che ne descrivono il contenuto. Il software impara quindi ad associare un'immagine alla corretta descrizione testuale e, data un'immagine casuale, a prevederne la descrizione corretta. Applicando infine il procedimento inverso, data una parola o una frase definita, la macchina sarà in grado di associare un'immagine pertinente alle descrizioni testuali fornite.

Caso studio e applicazioni

Il tema delle app *prompt-to-image* è stato proposto a studenti universitari nel laboratorio di *Pedagogia e didattica dell'arte* al secondo anno del corso di laurea in Scienze della formazione primaria.
Agli studenti è stato chiesto di progettare una narrazione visuale a partire da una fiaba o un racconto, utilizzando l'app Midjourney, a seguito di una fase di alfabetizzazione all'uso dell'applicazione e l'analisi di alcuni casi studio.
Agli studenti è stato chiesto di:

- immettere i comandi in inglese;
- partire sempre da una parola elementare, in grado di riassumere nel modo più aderente quello che si vuole generare;
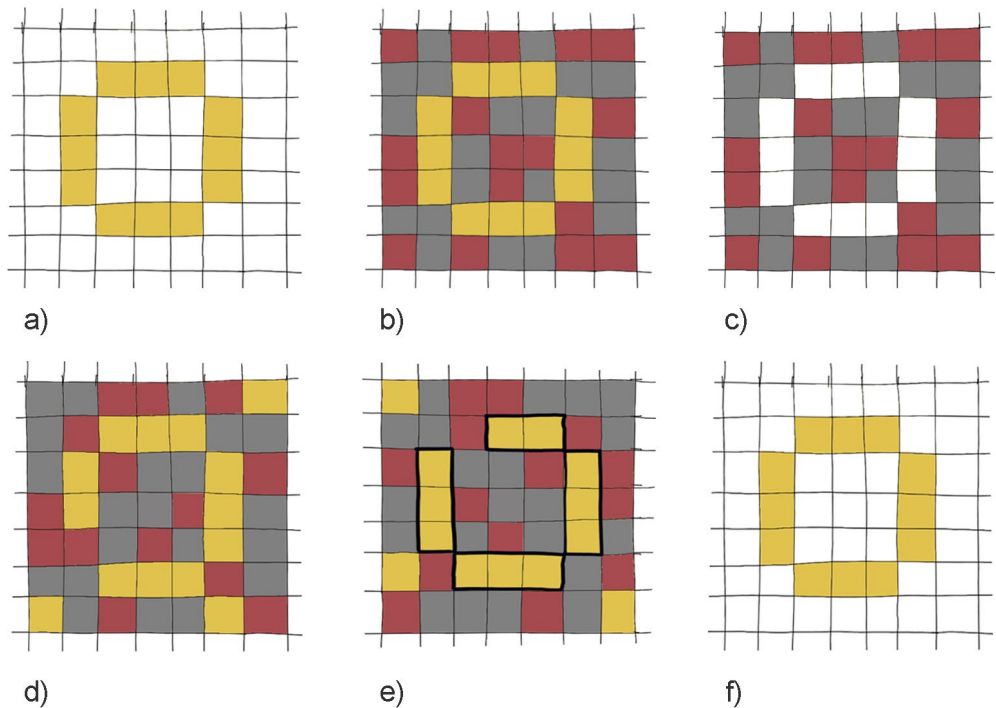
Fig. 6. Processo di generazione di immagini digitali nei sistemi *prompt-to-image*.

- specificare lo stile dell'immagine che si desidera generare. Aggiungere specifiche autoriali come 'una foto di' o 'un dipinto di';
- aggiungere qualunque parola specifica possa aiutare l'AI a capire in dettaglio alcune aspettative, come: dettagliato, realistico, etc.;

L'uso corretto del *prompting*, o come in alcuni contesti inizia ad essere definito *prompt design*, è divenuto un aspetto fondamentale dell'*imaging* ed è il mezzo con cui si esprime la creatività e la perizia tecnica dell'utente. Per questo è importante che gli utenti prima di immergersi nella produzione delle proprie immagini analizzino i *prompt* delle immagini rintracciabili in Discord, il *social network* in cui è possibile utilizzare Midjourney. Come in ogni processo di scrittura, a prescindere da quale sia il linguaggio, è fondamentale riconoscere i codici di base e saperli interpretare per poi poterli utilizzare per esprimere e comunicare il proprio messaggio. Si tratta quindi di una semiotica trasposta, dove il rapporto tra codice e immagine passa attraverso un linguaggio astratto come quello verbale.
Non sono state date indicazioni sullo stile delle immagini o sulla narrazione in maniera tale da far esplorare autonomamente tecniche e stili diversi e far emergere quanto più possibile l'originalità dell'elaborato. Per quanto, infatti, l'uso di un generatore automatico possa seguire procedure standardizzate per l'ottenimento dell'immagine e replicare uno stile autoriale specifico (cfr. fig. 2), allo stesso tempo permette di raggiungere un proprio stile personale tramite la scelta delle giuste parole chiave, ovviamente combinata con l'uso dei tool di modifica delle immagini che il software mette a disposizione.

Discussione

Gli studenti hanno realizzato i loro *visual storytelling* con una serie di 12 immagini e hanno prodotto degli elaborati che tenessero traccia dei *prompt* utilizzati. I risultati hanno dato esiti molto variabili, ma alcune considerazioni si possono generalizzare anche a seguito di una *survey* dei loro feedback a seguito della sperimentazione.
Innanzitutto, le apps *prompt-to-image* non sono ancora molto conosciute dagli studenti, e questo consente di elaborare riflessioni omogenee sui loro lavori in quanto realizzati in assenza di preconoscenze. Molti di loro riportano delle difficoltà nell'interpretare il rapporto tra *prompt*
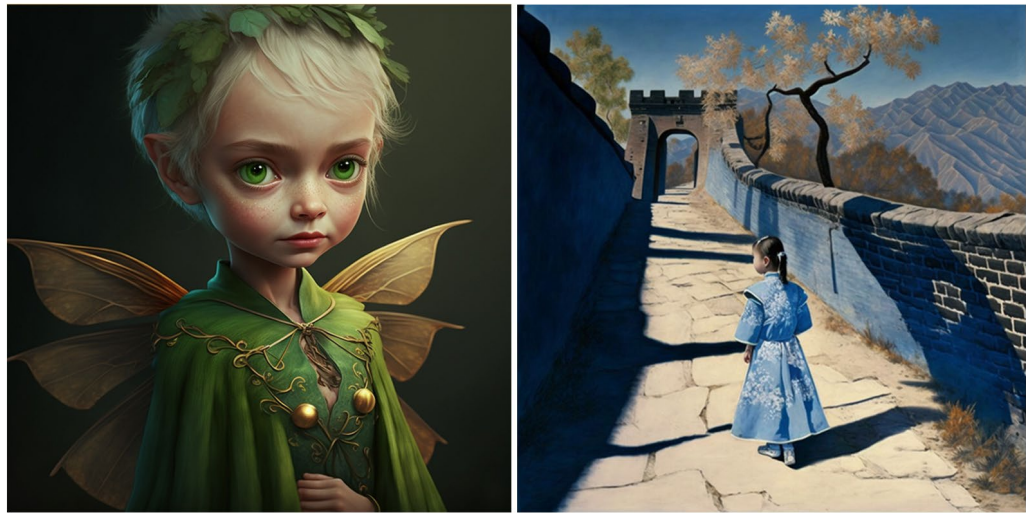
Fig. 7. Alcuni degli elaboratori prodotti dagli studenti impegnati nell'utilizzo di Midjourney.

e risultati attesi, sempre molto sorprendenti dal punto di vista dell'efficacia e del realismo, ma spesso lontani da una prefigurazione a livello mentale a cui gli studenti sembravano vincolati. In effetti è in questo passaggio del processo che risiede molta della attenzione che si deve dare alla generazione di immagini con motori AI, ma questo ci ha persuaso che ci sia, come anticipato, una necessità di educare alla lettura delle immagini per operare in modo disinvolto nella scrittura. I risultati, infatti, sono stati molto variabili ed è facile correlare alcune di queste differenze a differenti livelli di alfabetizzazione grafica e visuale (*graphic* and *visual literacy*). Alcuni studenti evidentemente hanno puntato tutta la loro attenzione alla coerenza delle immagini prodotte e sono stati soddisfatti dal loro realismo, come anticipato uno dei principali elementi di innovazione dell'automazione che questi sistemi ci consentono, senza cercare una scrittura scenografica autoriale, ovvero senza badare approfonditamente alle ambientazioni, allo stile visuale, ai dettagli o alle inquadrature: i risultati in questi casi sono meno convincenti, appiattiti su linguaggi retorici, più vicini all'illustrazione di servizio che a quella autoriale (fig. 6). L'autorialità, invece, in alcuni casi è stata più marcata, quando alcuni studenti hanno inserito nel *prompt* informazioni aggiuntive che consentissero di caratterizzare meglio i personaggi, le ambientazioni e lo stile visuale (fig. 7).

Conclusioni

Le possibilità di figurazione che questi strumenti ci offrono sono in continuità con i processi di automazione che nel campo della visualità e delle arti figurative sono da tempo in forte sviluppo. Gli elementi di vera innovazione sono un ulteriore upgrade del processo poietico, da sempre in parte demandato alla tecnologia utilizzata – dal pennello alla intelligenza artificiale –, e la disponibilità di un realismo che tali automazioni non avevamo mai fornito all'autore.
Disegnare con l'intelligenza artificiale significa, in definitiva, governare quel processo che parte dall'ideazione, si confronta con il mezzo tecnico e tramite i segni – i *prompt* che diventano tracciati – torna allo stato di elaborazione secondo un processo di circolarità ermeneutica – idea, segno, percezione, nuova ideazione – rendendo il processo di disegno e di ideazione inscindibili tanto con gli strumenti analogici che con quelli digitali.

Note

[1] Come un computer ha creato la copertina della settimana. *The Economist*, 11 giugno 2022.

[2] Per una lista di esempi, si veda <https://mpost.io/top-50-text-to-image-prompts-for-ai-art-generators-midjourney-and-dall-e/> (consultato il 30 gennaio 2023).

[3] Un esempio sono le illustrazioni create per la *Divina Commedia* di Dante. Tan 2022.

Riferimenti bibliografici

Alamar J. (Novembre 2022). The Illustrated Stable Diffusion. <https://jalammar.github.io/illustrated-stable-diffusion> (consultato il 15 gennaio 2023).

Carlson M. (6 settembre 2022). Stable Diffusion and why it matters. <https://hackaday.com/2022/09/06/stable-diffusion-and-why-it-matters/> (consultato il 30 gennaio 2023).

Choundary L. (23 agosto 2022). Stable Diffusion vs Midjourney vs DALL.E2. Text-to-image generators are gaining popularity this year. It began with DALL.E 2, but now we have amazing tools like Midjourney and Stable Diffusion and many more. In *AIM*. <https://analyticsindiamag.com/stable-diffusion-vs-midjourney-vs-dall-e2/> (consultato il 30 gennaio 2023).

Epstein Z., Levine S., Rand D.G. (2020). Who Gets Credits for AI-Generated Art? In *iScience,* vol. 23, 9. < https://www.sciencedirect.com/science/article/pii/S2589004220307070> (consultato il 15 gennaio 2023).

Growcoot M. (1 agosto 2022). Midjourney Generates AI Apocalyptic Images of the "Last Selfie Ever Taken". In PetaPixel. <https://petapixel.com/2022/08/01/dall-e-2-generates-ai-apocalyptic-images-of-the-last-selfie-ever-taken/> (consultato il 30 gennaio 2023).

Ho J., Jain A., Abbeel P. (2020). Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell (a cura di). *34th International Conference on Neural Information Processing Systems. Conference proceedings.* Vancouver, 6-12 dicembre 2020, pp. 6840-6851. New York: Curran Associates Inc.

How a computer designed this week's cover (11 giugno 2022). *The Economist*. <https://www.economist.com/news/2022/06/11/how-a-computer-designed-this-weeks-cover?> (consultato il 30 gennaio 2023).

Luigini A. (2019). Il foglio, lo schermo e lo stereoscopio digitale. Ovvero lo specchio, la retina invertita e la fine della rappresentazione. In P. Belardi (a cura di). *RIFLESSIONI l'arte del disegno / il disegno dell'arte*, pp. 137-142. Roma: Gangemi.

Miller A. (22 agosto 2022). MidJourney and the Advent of "Text to Museum Quality Art"? Examining the controversial AI generated art that is flooding your social streams. <https://oylmiller.substack.com/p/midjourney-and-the-advent-of-text> (consultato il 30 gennaio 2023).

Monge J.C. (6 settembre 2022). MidJourney VS Stable Diffusion: Same Prompt, Different Results. In *Medium*. <https://medium.com/codex/midjourney-vs-stable-diffusion-same-prompt-different-result-dd29ca482235> (consultato il 30 gennaio 2023).

Muppalla V., Hendryx S. (2022). Diffusion Models: A Pratical Guide. <https://scale.com/guides/diffusion-models-guide> (consultato il 15 gennaio 2023).

Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. (2022). High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition. Conference Proceedings*. New Orleans, 18-24 giugno 2022, pp. 10684-10695. <https://ieeexplore.ieee.org/document/9878449> (consultato il 30 gennaio 2023).

Sambucci L. (13 gennaio 2022). L'AI spiegata agli umani. DALL-E, cos'è e come funziona il sistema che genera immagini da testo. In *AI4Business*. <https://www.ai4business.it/intelligenza-artificiale/lai-spiegata-agli-umani-dall-e-il-sistema-che-genera-immagini-nuovo-alleato-di-creativi-e-designer/> (consultato il 30 gennaio 2023).

Sohl-Dickstein J., Weiss E.A., Maheswaranathan N., Ganguli S. (2015). Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In *ICML'15. 32th International Conference on Machine Learning. Conference proceedings*. Lille, Francia, 6-11 Luglio 2015, vol. 37 pp. 2256-2265.

Tan B. (11 agosto 2022). Dante's Midjourney, and our AI Future. <https://www.bryanjtan.com/post/dante-midjourney-and-our-ai-generated-future> (consultato il 30 gennaio 2023).

Vincent J. (2 agosto 2022). "An engine for the imagination": the rise of AI image generators. An interview with Midjourney founder David Holz. In *The Verge*. <https://www.theverge.com/2022/8/2/23287173/ai-image-generation-art-midjourney-multiverse-interview-david-holz> (consultato il 30 gennaio 2023).

Zylinsk J. (2020). *AI - Art Machine Visions and Warped dreams.* London: Open Humanities Press.

Autori
Francesca Condorelli, Libera Università di Bolzano, francesca.condorelli@unibz.it
Alessandro Luigini, Libera Università di Bolzano, alessandro.luigini@unibz.it
Giuseppe Nicastro, Libera Università di Bolzano, giuseppe.nicastro@unibz.it
Barbara Tramelli, Libera Università di Bolzano, barbara.tramelli@unibz.it

Per citare questo capitolo: Condorelli Francesca, Luigini Alessandro, Nicastro Giuseppe, Tramelli Barbara (2023). Disegno e intelligenza artificiale. Enunciati teorici e prassi sperimentale per una poiesi condivisa/Drawing and Artificial Intelligence. Theoretical Statements and Experimental Practice for a Shared Poiesis. In Cannella M., Garozzo A., Morena S. (a cura di). *Transizioni. Atti del 44° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione/Transitions. Proceedings of the 44th International Conference of Representation Disciplines Teachers*. Milano: FrancoAngeli, pp. 2623-2640.

Isbn 9788835155119

Drawing and Artificial Intelligence. Theoretical Statements and Experimental Practice for a Shared Poiesis

Francesca Condorelli
Alessandro Luigini
Giuseppe Nicastro
Barbara Tramelli

Abstract

The pervasiveness of digital technologies is fostering a rapid development of the set of technologies that fall within the Artificial Intalligence (AI) designation, and how many of the digital technological innovations have a major impact on the production of images and drawings. Studying how creativity interfaces with an AI system endowed with considerable autonomy in the production of an image is critical to understanding how the author must operate to govern the tool and guide its poiesis. The paper proposes a theoretical contextualization of the phenomenon of text-to-image applications in the landscape of art history and image production, closely linked with an explication of how such applications work, and then verifies their potential in the field of visual storytelling.

Keywords
Artificial intelligence, Visual storytelling, Poiesis

Some examples of images generated with the prompt-to-image application Midjourney.

doi.org/10.3280/oa-1016-c426

Introduction

The pervasiveness of digital technologies is fostering a rapid development of the set of technologies that fall within the designation Artificial Intelligence (AI), with web-based applications accessible from any computing device, including mobile devices. The generation of verbal and visual text has become possible thanks to AI solutions such as ChatGPT and Midjourney, and the results are convincing, though still in the process of further development. The paper presented here starts with a theoretical statement of the context in which these applications are developed, continues with a description of the process by which the images are generated, and ends with the presentation of an experiment in the educational setting used as a pragmatic verification of the previously stated theses.

Theoretical reflection

The models of AI which can create images from prompt texts (the so-called text-to-image generators) such as Midjourney, DALL.E and Stable Diffusion are increasingly becoming popular tools for artists and designers, and as it happens in many endeavors concerning the growing field of the digital humanities, theory necessarily comes after practice. The phenomenon of machine learning systems creating images from database sets has become worldwide known, and questions related to their use and usability (for instance, about copyright) are already been dealt with [1]. Their increasing popularity calls for the need to give them a theoretical frame and context in the evolving field of visual imaging.
Such tools allow a descriptive representation based on the keywords inserted, and the results can be surprisingly detailed in terms of visual outcomes [2]. As for matters of image quality, Midjourney especially tends to have a specific detailed style which makes it recognizable, while others (e.g. DALL.E) have the tendency to forget keywords when they become too many [Sambucci 2022]. The style itself is quite consistent in different images, as the founder David Holz stated: "we are focused toward making everything beautiful and artistic looking" and that is one striking characteristic about the images created: they generally look appealing to the eye, a sort of 'eye-candy' for the beholder [Vincent 2022]. The quality of these images can be interpreted considering both the level of realism in the visual output as well as the number of details which faithfully represent and interpret the linguist prompt inserted. The second point is particularly tantalizing to consider, as it entails a closer analysis of the relationship between human and computer interaction, which does not limit itself to 'put a word in, get an image out'. The process of creating these visual outputs is much more complex than that, and it shows especially when the system 'fails' to generate the expected result, such as different attempts can show [Carlson 2022].
Actually, even Midjourney occasionally omits some details inserted in the prompts, while maintaining a convincing level of imaging (fig. 1). The two main problems encountered during the use of the algorithm are a general intrinsic difficulty in managing the typographic aspects (fig. 2), and the inability – difficult to understand at times, when compared to the realism of faces and other generated details – in the generation of hands, both in terms of shape and anatomical correctness (fig. 3).
As is well known, nowadays AI learns through trial and error, based on data we feed it. The more data, the more accurate it usually becomes. The results can be much more or much less than expected, as the popularity of the generated images after their creation. Midjourney for instance usually offers dreamlike and often nightmarish images with a 'neo-gothic' style, both appealing to the eye and intuitive to interpret, which makes the model particularly apt to illustrate novels, stories, and fairytales [3]. The role of the AI is central to the process, but of course the role of the user who inserts the keywords is equally important. As stated in a recent article by Ziv Epstein: "No AI acts alone, completely divorced from the influence of humans. Even the artwork *Edmond de Belamy*, which was claimed to be 'entirely realised by an algorithm' was the result of the creativity, hard work, and decisions of numerous human contributors. When an AI system achieves something great or causes a serious

Fig. 1. #*Fallingwater house by Frank Lloyd Wright on mars with a blue sunset and a flying horse*. Evidently, Midjourney's elaboration lacks the winged horse, although it has generated a visually striking image with its own semantic coherence and correspondence with the prompt.

Fig. 2. #*happy new year, 2023, Yayoi Kusama style*. In this quatern of Midjourney-generated images, the difficulty of handling typefaces is evident: the AI understands that it is necessary to insert a number into the image, probably from the first part of the prompt 'happy new year' but fails to insert the correct number, despite the fact that it is present in the prompt and has been isolated from punctuation. The artist's style, however, is correctly interpreted.

Fig. 3. Left, #*drawing resembling a black and white engraving of a hands that old a pencil. A destra, #black and white illustration of a hand.* These series of images are two of many elaborate series on the theme of drawing hands, and the results, as we can clearly see, are very fascinating but completely lose the references of anatomical correctness and verisimilitude that make Midjourney's images so effective.

problem, how is responsibility attributed to the humans surrounding it?" [Epstein et al. 2020]. In the case of Midjourney, we see the essential role played by keywords, and therefore by the users. The process of creating art from analyzing artistic material (in this case millions of web images) and amalgamating this material without authoritative control is not new in art, as creative experiments of the like were done well before computer generated art (insert examples – performances, dripping, etc.).
The weight and role of technique and technology in artistic production has always been a vital element for distinguishing styles, periods, intentions of individual authors and systems of collaboration between multiple authors. From the moment in which a certain technology can be used by several people (not necessarily simultaneously or in the same way), the question of authorship – widely debated in the history of art – is and remains open.
When in 1975 Brian Eno publishes the concept album *Discreet Music*, in which he experiments with different sound generation techniques instead to traditional musical instruments, he explicitly declares his poietic program on the album cover: "Since I have always preferred making plans to executing them, I have gravitated towards situations and systems that, once set into operation, could create music with little or no intervention on my part. That is to say, I tend towards the roles of planner and programmer, and then become audience to the results".
Countless experiences could be cited in the field of visual aesthetic expressions, but for the sake of brevity it will be sufficient to recall that since the 1960s experiments – such as that of Desmond Paul Henry and his *Drawing Machines* – electro-mechanical devices capable of tracing signs on a sheet following mathematical patterns – the attempts to pursue an automated poietic process diversified -with mixed results.
This is the case for Adrian Ward's *Autoillustrator* or Oliver Auber's *Poietic Generator* which investigate the possibilities of a collaborative and non-automated *poiesis*: the users-authors are the ones who, by interacting, generate visual configurations, thus creating a shared poiesis which is made possible by idea of the system's author. Our excursus could continue by quoting the pioneers of digital architecture who in the 1990s investigated the potential of advanced modelers and generative systems for non-traditional poietic paths, such as Marcos Novak, Greg Lynn, Karl Chu or the Asymptotes (fig. 4).
From this point of view it is not the automation (the image generation system distinct from the direct action of the author) that marks the difference between the prompt-to-image AI systems and the previous technical and technological devices. Among all the differences we wish to underline the most evident one: all the works mentioned above, equally self-generated, are abstract aesthetic expressions, while Midjourney or Dall-E allow us to access a previously inaccessible and perfectly photorealistic imaginary. The amazement that the first approach to these systems produces in the generic user is usually given by the realism, by
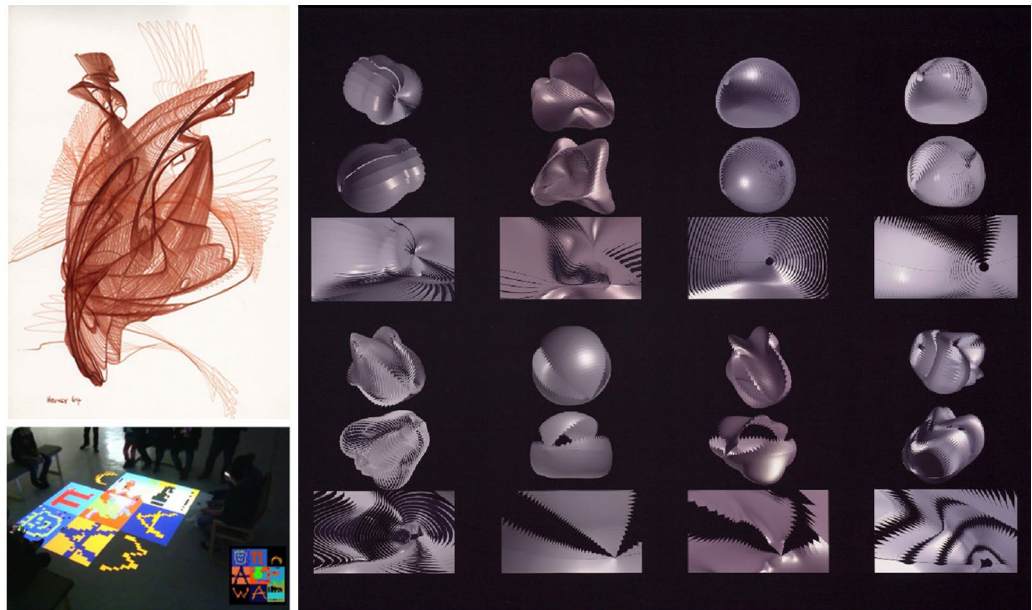
Fig. 4. Top left, Desmond Paul Henry, *Untitled*, 1964, <desmondhenry.com>. Bottom left, Olivier Auber, *Poietic Generator*, 1986-2017, <narrabase.net/auber.html>. Right, Greg Lynn, *Embryologic House*, 1998, <sfmoma.org/artist/Greg_Lynn>.

the anatomical and physiognomic coherence, both anthropomorphic and zoomorphic, and by the lighting, which demonstrate how these systems are able to conceive the relationship between the parts of an image by representing its three-dimensionality without simulation. The convincing realism opens up many imaging applications and the application possibilities are endless. There are already several cases of prizes won by artists who presented a work created in Midjourney, such as Jason Allen, or by creatives who invented sets, backdrops or entire series of characters in Midjourney, such as the *Bear Verse* videogame by Eugene Kitkin, but the authorship does not seem to be doubtful – for now.

Image Creation systems based on Artificial Intelligence: a general overview

The rise of AI solutions capable of generating images from textual prompts cannot be attributed to a single line of study, but rather to the convergence of different research that made possible the birth of applications such as Midjourney, Dall-E, etc. [Muppalla et al. 2022].

The process consists of three steps (fig. 5): first, a text command is sent to a text encoder that has been trained to decipher it and assign numerical values to it. The second step involves the intervention of a model called 'prior', which associates the encoding of the text with an encoding of the corresponding image, acquiring the semantic information of the text command. Finally, an 'image decoder' generates an image that is a visual manifestation of the same semantic information.

Relative to the way in which images take shape, this stage of the process is related, from a technical point of view, to the possibility of deconstructing an organised set of specific data and then applying a reverse process that is able to reconfigure the data into a composition as close as possible to the original one [Sohl-Dickstein et al. 2015]. Relating to digital images, it is therefore a matter of altering the pixels that constitute it, whose precise position within the digital space gives life to what we see (fig. 6a), by introducing a known amount of noise (i.e. a quantity of randomly positioned pixels) that alters the image itself (figs. 6b, 6c). The next stage of this process therefore involves acting in reverse (figs. 6d, 6f) to reconstruct the original position of the pixels, thus resulting in a new image similar to the initial one [Rombach et al. 2022, Ho et al. 2020].

By reiterating this process on large datasets, it is then possible to generate a predictive model that can independently compose images by reordering a quantity of unstructured pixels.
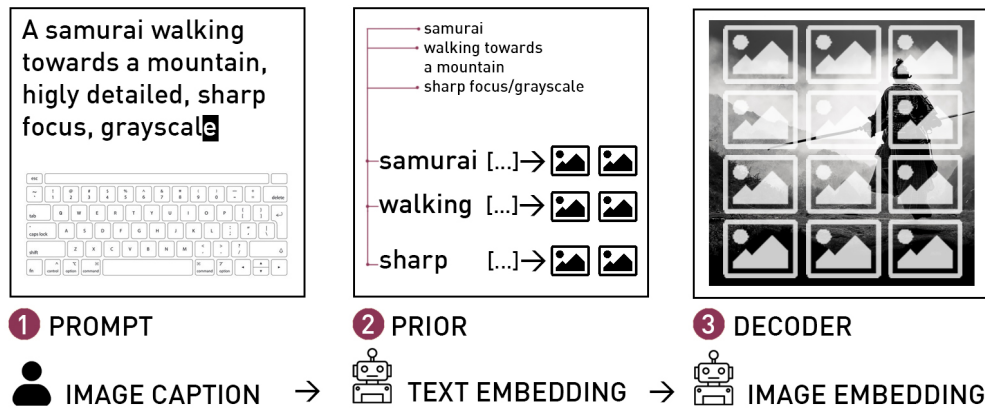
Fig. 5. Synthetic workflow of a prompt-to-image system.

What has been described so far thus allows us to generate images randomly: the next step is therefore represented by the possibility of inserting, within this chain, a series of instructions that can guide the generation process towards a predetermined and non-random direction.
The possibility of being able to teach a computer how to discriminate and classify the content of an image is certainly one of the fields of research that has most interested the scientific community dealing with Machine Learning and Artificial Intelligence. Generate an image from a textual prompt requires the machine to first of all be able to interpret a graphic object (a photo, a drawing, etc.) and associate it with a textual tag describing its content [Alamar 2022]. Also in this case, the training phases involve the analysis of datasets consisting of millions of images associated with their respective caption describing their content. The software then learns to associate an image with the correct textual description and, starting from a random image, to predict its correct description. Finally, by applying the reverse procedure, given a defined word or phrase, the machine will be able to associate a relevant image with the given text descriptions.

Case Studies and Applications

The topic of prompt-to-image apps was proposed to university students in the Art Pedagogy and Didactics laboratory in the second year of the Primary Education degree course. Students were asked to design a visual narrative from a fairy tale or short story, using the Midjourney app, following a literacy phase in the use of the app and the analysis of some case studies. Students were asked to:

- enter commands in English;
- always start with an elementary word that most closely sums up what they want to generate;
- specify the style of the image to be generated. Add specifications such as 'a photo of' or 'a painting of';
- add any specific words that may help the AI understand certain expectations in detail, such as: detailed, realistic, etc.;

The proper use of prompting, or as it is beginning to be called prompt design in some contexts, has become a fundamental aspect of imaging and is the means by which the user's creativity and technical expertise is expressed. This is why it is important for users to analyse image prompts that can be found in Discord, the social network where Midjourney can be used, before diving into the production of their own images. As in any writing process, no matter what the language, it is essential to recognise the basic codes and know how to interpret them in order to then be able to use them to express and communicate one's message. It is therefore a transposed semiotics, where the relationship between code and image
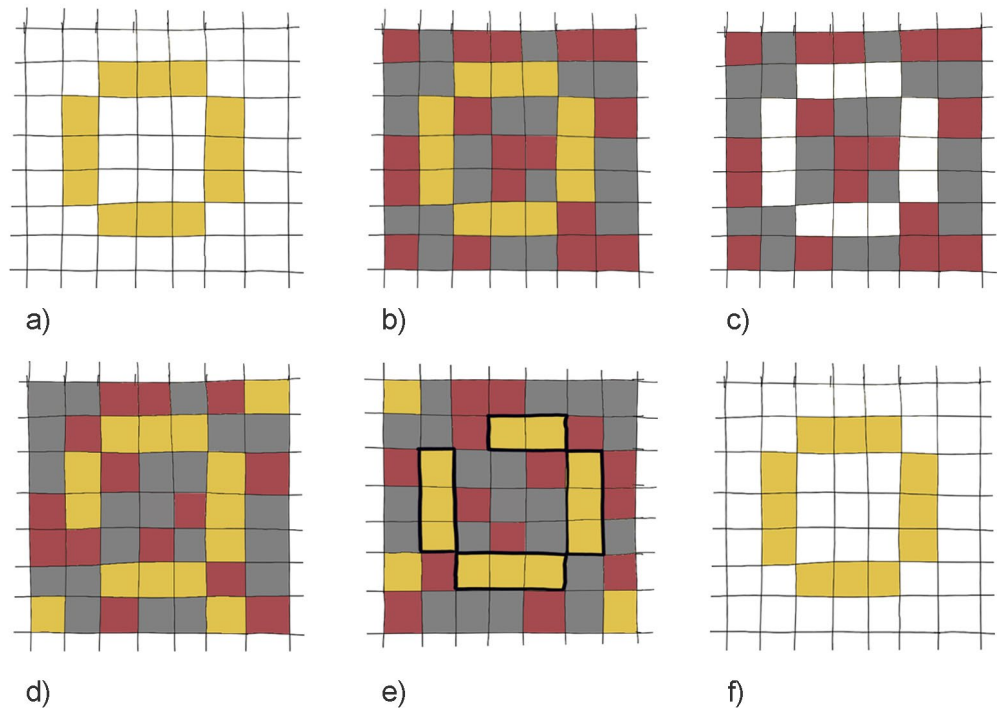
Fig. 6. Process of digital image generation in prompt-to-image systems.

passes through an abstract language such as the verbal language. No indications were given as to the style of the images or the narration so that different techniques and styles could be explored independently, and the originality of the work could emerge as much as possible. In fact, although the use of an automatic generator can follow standardised procedures for obtaining the image and replicate the style of famous artists, at the same time it allows one to achieve one's own personal style through the choice of the right keywords, obviously combined with the use of the image editing tools that the software makes available.

Discussion

The students realised their visual storytelling with a series of 12 images and produced illustration keeping track of the prompt used. The results are highly variable, but some considerations can also be generalised from a survey of their feedback following the experimentation.

First of all, prompt-to-image apps are not yet very well known to the students, which allows for homogeneous reflections on their work as they have no prior knowledge of them. Many of them report difficulties in interpreting the relationship between prompts and expected results, which are always very surprising from the point of view of effectiveness and realism, but often far from a prefiguration to which the students seemed to be attached. Indeed, it is in this step of the process that much of the attention that needs to be given to image generation with AI engines resides, but this persuaded us that there is, as anticipated, a need to educate in image reading in order to operate fluently in writing. The results, in fact, were highly variable and it is easy to relate some of these differences to different levels of visual literacy. Some students evidently focused all their attention on the coherence of the image produced and were satisfied with their realism, as anticipated one of the main elements of innovation in automation that these systems allow us, without seeking real scenographic writing, i.e. without paying much attention to settings, visual style, details or framing: the results in these cases are less convincing, flattened on rhetorical language, closer to service illustration than to authorship. Authorship, on the other hand, was more pronounced in some cases, when some students included additional information in the prompt to better characterise the characters, the settings and the visual style (fig. 7).

Fig. 7. Some of the processors produced by students using Midjourney.

Conclusions

The possibilities of figuration that these tools offer us are in continuity with the processes of automation that have been developing strongly in the field of visual and figurative arts for some time. The truly innovative elements are a further upgrade of the poietic process, which has always been partly entrusted to the technology used – from the paintbrush to artificial intelligence – and the availability of a realism that such automations had never provided to the author.
Drawing with artificial intelligence means, definitively, governing that process that starts from ideation, is confronted with the technical medium and through the signs drawn – the prompts – returns to the state of elaboration according to a process of hermeneutic circularity – idea, sign, perception, new ideation – making the process of drawing and ideation inseparable from both analogue and digital tools.

Notes

[1] How a computer designed this week's cover. *The Economist*, 11 June 2022.

[2] For a list of samples, see for instance: <https://mpost.io/top-50-text-to-image-prompts-for-ai-art-generators-midjourney-and-dall-e/ > (accessed 30 January 2023).

[3] One example can be the illustrations created for Dante's *Divina Commedia*. Tan 2022.

References

Alamar J. (November 2022). The Illustrated Stable Diffusion. <https://jalammar.github.io/illustrated-stable-diffusion> (accessed 15 January 2023).

Carlson M. (6 September 2022). Stable Diffusion and why it matters. <https://hackaday.com/2022/09/06/stable-diffusion-and-why-it-matters/> (accessed 30 January 2023).

Choundary L. (23 August 2022). Stable Diffusion vs Midjourney vs DALL.E2. Text-to-image generators are gaining popularity this year. It began with DALL.E 2, but now we have amazing tools like Midjourney and Stable Diffusion and many more. In *AIM*. <https://analyticsindiamag.com/stable-diffusion-vs-midjourney-vs-dall-e2/> (accessed 30 January 2023).

Epstein Z., Levine S., Rand D.G. (2020). Who Gets Credits for AI-Generated Art? In *iScience*, Vol. 23, 9, pp. 1-4. < https://www.sciencedirect.com/science/article/pii/S2589004220307070> (accessed 15 January 2023).

Growcoot M. (1 agosto 2022). Midjourney Generates AI Apocalyptic Images of the "Last Selfie Ever Taken". In *PetaPixel*. <https://petapixel.com/2022/08/01/dall-e-2-generates-ai-apocalyptic-images-of-the-last-selfie-ever-taken/> (consultato il 30 gennaio 2023).

Ho J., Jain A., Abbeel P. (2020). Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell (Eds.). In *34th International Conference on Neural Information Processing Systems. Conference proceedings*. Vancouver. 6-12 December 2020, pp. 6840-6851. New York: Curran Associates Inc.

How a computer designed this week's cover (11 June 2022). *The Economist*. <https://www.economist.com/news/2022/06/11/how-a-computer-designed-this-weeks-cover?> (accessed 30 January 2023).

Luigini A. (2019). The sheet, the screen and the digital stereoscope. Or the mirror, the inverted retina and the end of the representation. In P. Belardi (Ed.). *RIFLESSIONI l'arte del disegno/il disegno dell'arte*, pp. 137-142. Rome: Gangemi.

Miller A. (22 August 2022). MidJourney and the Advent of "Text to Museum Quality Art"? Examining the controversial AI generated art that is flooding your social streams. <https://oylmiller.substack.com/p/midjourney-and-the-advent-of-text> (accessed 30 January 2023).

Monge J.C. (6 September 2022). MidJourney VS Stable Diffusion: Same Prompt, Different Results. In *Medium*. <https://medium.com/codex/midjourney-vs-stable-diffusion-same-prompt-different-result-dd29ca482235> (accessed 30 January 2023).

Muppalla V., Hendryx S. (2022). Diffusion Models: A Pratical Guide. <https://scale.com/guides/diffusion-models-guide> (accessed 15 January 2023).

Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. (2022). High-resolution image synthesis with latent diffusion models. *IEEE/CVF Conference on Computer Vision and Pattern Recognition. Conference proceedings*. New Orleans, 18-24 June 2022, pp. 10684-10695. <https://ieeexplore.ieee.org/document/9878449> (accessed 30 January 2023).

Sambucci L. (13 January 2022). L'AI spiegata agli umani. DALL-E, cos'è e come funziona il sistema che genera immagini da testo. In *AI4Business*. <https://www.ai4business.it/intelligenza-artificiale/lai-spiegata-agli-umani-dall-e-il-sistema-che-genera-immagini-nuovo-alleato-di-creativi-e-designer/> (accessed 30 January 2023).

Sohl-Dickstein J., Weiss E.A., Maheswaranathan N., Ganguli S. (2015). Deep Unsupervised Learning using Nonequilibrium. In *ICML'15. 32th International Conference on Machine Learning. Conference proceedings*. Lille, France, 6-11 July 2015, Vol. 37 pp. 2256-2265.

Tan B. (11 August 2022). Dante's Midjourney, and our AI Future. <https://www.bryanjtan.com/post/dante-midjourney-and-our-ai-generated-future> (-il-sistema-che-genera-immagini-nuovo-alleato-di-creativi-e-designer/> (accessed 30 January 2023).

Vincent J. (2 August 2022). "An engine for the imagination": the rise of AI image generators. An interview with Midjourney founder David Holz. In *The Verge*. <https://www.theverge.com/2022/8/2/23287173/ai-image-generation-art-midjourney-multiverse-interview-david-holz> (accessed 30 January 2023).

Zylinsk J. (2020). *AI - Art Machine Visions and Warped dreams*. London: Open Humanities Press.

Authors
Francesca Condorelli, Libera Università di Bolzano, francesca.condorelli@unibz.it
Alessandro Luigini, Libera Università di Bolzano, alessandro.luigini@unibz.it
Giuseppe Nicastro, Libera Università di Bolzano, giuseppe.nicastro@unibz.it
Barbara Tramelli, Libera Università di Bolzano, barbara.tramelli@unibz.it

To cite this chapter: Condorelli Francesca, Luigini Alessandro, Nicastro Giuseppe, Tramelli Barbara (2023). Disegno e intelligenza artificiale. Enunciati teorici e prassi sperimentale per una poiesi condivisa/Drawing and Artificial Intelligence. Theoretical Statements and Experimental Practice for a Shared Poiesis. In Cannella M., Garozzo A., Morena S. (Eds.). *Transizioni. Atti del 44° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione/Transitions. Proceedings of the 44th International Conference of Representation Disciplines Teachers*. Milano: FrancoAngeli, pp. 2623-2640.

Isbn 9788835155119